Effect of Image Captioning with Description on the Working Memory

by

Nithiya Shree Uppara

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved July 2021 by the
Graduate Supervisory Committee:

Troy McDaniel, Co-Chair
Hemanth Venkateswara, Co-Chair
Chris Bryan

ARIZONA STATE UNIVERSITY

August 2021

ABSTRACT

Working memory plays an important role in human activities across academic, professional, and social settings. Working memory is defined as the memory extensively involved in goal-directed behaviors in which information must be retained and manipulated to ensure successful task execution. The aim of this research is to understand the effect of image captioning with image description on an individual's working memory. A study was conducted with eight neutral images comprising situations relatable to daily life such that each image could have a positive or negative description associated with the outcome of the situation in the image. The study consisted of three rounds where the first and second round involved two parts and the third round consisted of one part. The image was captioned a total of five times across the entire study. The findings highlighted that only 25% of participants were able to recall the captions which they captioned for an image after a span of 9-15 days; when comparing the recall rate of the captions, 50% of participants were able to recall the image caption from the previous round in the present round; and out of the positive and negative description associated with the image, 65% of participants recalled the former description rather than the latter. The conclusions drawn from the study are participants tend to retain information for longer periods than the expected duration for working memory, which may be because participants were able to relate the images with their everyday life situations and given a situation with positive and negative information, the human brain is aligned towards positive information over negative information.

DEDICATION

Dedicated to my late grand mother, Smt. Venkateshwaramma Askani.

ACKNOWLEDGEMENTS

The submission of my thesis concludes my incredible graduate journey at Arizona State University (ASU). I would like to express my gratitude to many wonderful people who have helped me reach this place.

First and foremost, I would like to sincerely thank my advisors, Dr. Hemanth Venkateswara and Dr. Troy McDaniel, for their continuous support and encouragement. Without their guidance and motivation, my thesis would not have been possible and complete. I enjoyed an enormous amount of freedom in working on the problems of my interest. I was also constantly supported up financially and emotionally by Dr. Hemanth and Dr. Troy. I would also like to thank my committee member, Dr. Chris Bryan, for his support, guidance, and inspiring conversation.

I am blessed to have had the opportunity to work with an amazing set of peers at CUbic Lab. I would like to thank Badri and Raghav for their support and for having many productive conversations. I would also like to thank Ph.D. students Bijan and Meredith for sharing their incredible experiences. My time at ASU would not have been more exciting without the love and support of my friends Srinivas, Venu Gopinath, Srushti, Yasaswini, Manasa, Pavan and Chandrakaran.

Finally, my thesis would not have been completed successfully without my parents, Dr. Jaya Laxmi and Srinivasulu, my younger brother, Ruthvik Sai, my grandparents, Satyanarayana and Venkateshwaramma, and my uncle, Venugopal. Lastly, I thank my husband Rajashekar Reddy for being constant support and inspiration. He has always inspired me to push my limits further.

TABLE OF CONTENTS

Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 Goals and Motivations

Working memory has been a fascinating area of research since its introduction in the 1960s Baddeley (2012); D'Esposito and Postle (2015). Various studies about memory in the fields of psychology, biology, or neuroscience have not been able to completely outline a categorization of memory in terms of its functionality and mechanism Baddeley (2010); Cowan (2008); Oberauer and Cowan (2005). Working memory has been gaining a lot of importance in mundane human activities, such as academic, professional and social settings Harden (2011).

The goal of my dissertation is to understand the effect of image captioning with description on the working memory of humans. It tries to understand the importance of working memory on human behavior in day-to-day activities such as image captioning.

This dissertation has been inspired by some overarching challenges and goals in psychology and computer science. This research plans to design, develop and analyze an image captioning game and explore various nuances in the area of working memory of humans.

1.2 Contributions

The contributions of the dissertation are as follows.

1. Design a novel game which aims to understand the effect of image captioning with different descriptions i.e, positive and negative description, on the working

memory of the human being.

2. Collect neutral images and generate positive and negative descriptions corresponding to the outcome of the neutral images.

3. Develop the game with different levels using the neutral images and their corresponding image descriptions.

4. Perform human subjects research and collect data. Analyze the data to get different insights from the data.

5. Suggest future enhancements to the game to further study the reasons behind the effect of image captioning on working memory

1.3 Dissertation Outline

The dissertation is structured in the following manner.

Chapter 2 is a literature survey on working memory. Firstly, the survey provides an introduction to memory and the differences between the types of memories a human possesses. It provides more details about working memory and the factors affecting the capacity of working memory. It further outlines various studies related to my problem statement and corresponding outcomes of the studies. Finally, the chapter concludes by providing an overview of the study and the hypothesis related to it.

Chapter 3 describes the image dataset designed for the experimental study. The chapter provides details about all the images and their corresponding description used in the study.

Chapter 4 describes the experimental study in detail. The first section provides information regarding the participants of the study. The second section outlines the details of the study. It further elaborates the game design and procedure of the study. The third part provides the analysis of the data collected in the study. It also talks about the methods used to analyze the data. This chapter also contains images of interfaces used in the image captioning game for various rounds.

Chapter 5 provides the results generated after analyzing the data from the image captioning study. The first section shows the distribution of image captions, considering the context similarity between Round-1 Part-1 and Round-3 captions of the study. The second section describes the distribution of image captions, considering the context similarity between Round-1 and Round-2 Part-2, and considering the context similarity between Round-2 and Round-3 of the study. The final section provides the impact of the description of images on Round-1 Part-2 and Round-2 Part-2 captions.

Chapter 6 discusses the results from the previous chapter. It further makes a detailed analysis of the results and the reasons behind the obtained results. It also relates the results with the previous researches and provides critical analysis.

Chapter 7 describes the possible limitations of the current study. It also accounts for the various assumptions and interpretations considered during the study.

Chapter 8 concludes the dissertation by summarizing the contributions of the image captioning study. It further outlines directions for future research in this area. It provides possible enhancements and improvements to the current study.

Chapter 2

LITERATURE SURVEY

2.1 Memory

The quest to understand the human brain and the workings of human memory has intrigued philosophers and researchers for centuries. Memory is one of the most important aspects of what makes us human, and yet it is one of the most elusive and misunderstood of human faculties. Memory can be pictured as a small filing cabinet with separate memory folders where information is kept, or as a brain supercomputer with enormous capacity and speed MacDonald (2008). To retrieve a memory from the past, different areas of the brain collaborate. For example, let's consider the act of driving a car which is recreated by the brain from many different areas: the memory of how to get from the current location to the end of the block, the memory of how to operate the car, and the memory of driving the car while following the safety rules, which all come from different parts of the brain. Each memory element (sights, sounds, phrases, and emotions) is encoded in the same portion of the brain that created that fragment in the first place, and recalling a memory effectively reactivates the neural patterns that were established during the original encoding MacDonald (2008). A lasting memory in the brain is created when all the different types of memory work together to form it. The popular Atkinson-Shiffrin model defines a 3-step model for memory including sensory memory, short-term memory or working memory, and long-term memory Atkinson and Shiffrin (1968).

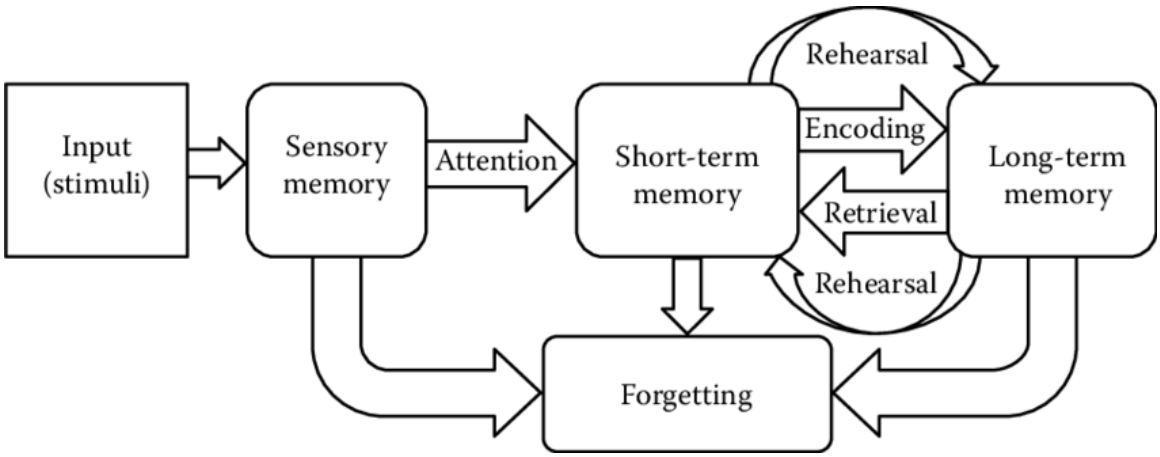


Figure 2.1: Atkinson And Shiffrin Memory Model

2.2 Long-term Memory and Short-term Memory

To understand the basic definition of working memory, one must first understand the difference between long-term memory and short-term memory. Long-term memory is defined as a vast store of knowledge and a record of prior events Cowan (2008). Long-term memory capacity varies from situation to situation and from person to person. Short-term memory is the ability of the human mind to hold a finite amount of information in a very accessible state, temporarily Atkinson and Shiffrin (1968). The main difference between long-term memory and short-term memory is the duration of the situation of information stored and the capacity of the information stored Broadbent (1958). The former has a huge capacity to retain information for a long duration and the latter is limited by the total number of chunks of information that can be stored at a time Cowan (2008).

2.3 Working Memory

Working memory is not completely different from short-term memory. Working memory is defined as the memory extensively involved in goal-directed behaviors in which information must be retained and manipulated to ensure successful task execution Chai *et al.* (2018). Miller *et al.* (1960) proposed the term working memory to refer to memory as it is used to plan and carry out the behavior. An example of common use of working memory is recalling partial calculations while solving a mathematical problem. The information stored during this process is stored only for that instance of time and is discarded from memory when the purpose is served. The factors related to the amount of time the information is stored change depending on the situation in which the information is perceived. Working memory assessments have been found to correlate with intellectual aptitudes (particularly fluid intelligence) better than short-term memory measures, and possibly better than assessments of any other psychological process Oberauer and Cowan (2005); Daneman and Carpenter (1980); Kyllonen and Christal (1990); Daneman and Merikle (1996); Engle *et al.* (1999b).

The multi-component working memory model, which is one of the most often mentioned working memory models in literature Oberauer and Cowan (2005); Baars and Franklin (2003); Ashkenazi *et al.* (2013); D'Esposito and Postle (2015); Kim *et al.* (2015), was the first to characterize working memory.. This model states that working memory manipulates information storage for greater and more complex cognitive utility Baddeley and Hitch (1974); Davelaar (2013). Working memory in this model consists of three basic stores: the central executive, the visuospatial sketchpad, and the phonological loop Baddeley and Hitch (1974). The central executive essentially

serves as a sensory store for attention. It sends data to the phonological loop, the visuospatial sketchpad, and the episodic buffer, which are all part of the same system. The articulatory process uses the phonological loop to store auditory information by silently rehearsing sounds or words in a continuous loop. The visuospatial sketchpad stores visual and spatial information that comes into the picture when performing spatial tasks such as judging speed or visual ones such as counting the doors on a house. The episodic buffer is used to link information between domains to form integrated units of visual, spatial, and verbal information and chronological ordering. The memory of a story or a book is an example of chronological information stored in an episodic buffer. Long-term memory and semantic meaning are also assumed to be linked to the episodic buffer.

I would like to refer to the definition of working memory as the ability to maintain goal-relevant information in mind for my further details in this paper. It is believed by many researchers Engle *et al.* (1999a); Just and Carpenter (2018); Kyllonen and Christal (1990); Daneman and Carpenter (1980) that working memory is fundamental to higher cognitive functions, including reasoning and reading comprehension, and is also linked to scholastic development Hitch *et al.* (2001). The amount and nature of information maintained in the working memory depend on the pattern and magnitude of brain activation during the working memory task Thomason *et al.* (2009). Many researchers Kirschen *et al.* (2005); Zarahn *et al.* (2004); Jaeggi *et al.* (2003); Veltman *et al.* (2003); Rypma *et al.* (1999) have proposed that maintaining larger amounts of information in working memory leads to higher activation during working memory tasks until the capacity limitations are reached Jansma *et al.* (2004); Callicott *et al.* (1999). A lot of behavioral studies Conklin *et al.* (2007); Gathercole *et al.* (2004); Towse *et al.* (2001); Chelonis *et al.* (2000) have shown that there is an increase in

working memory ability from childhood to adulthood.

The ability to maintain task-relevant visual information in the absence of external input (i.e., visual short-term memory) is critical for learning new skills, solving novel tasks, and acquiring new knowledge from the ever-changing visual world Logie and Logie (1995). The most important characteristic of visual working memory is its inherently capacity-limited nature. Many researchers Alvarez and Cavanagh (2004); Cowan (2001); Luck and Vogel (1997) stated that the working memory capacity of an individual is approximately three to four. The importance of working memory in acquiring knowledge from the visual environment, in conjunction with its capacity-limited nature, highlights the need for investigating whether it is possible to enhance it via training. Working memory plays a critical role in guiding these top-down attentional processes by keeping current goals in mind Downing (2000); de Fockert *et al.* (2001); Lavie *et al.* (2004). de Fockert *et al.* (2001), Lavie *et al.* (2004) and Lavie (2005) have demonstrated that increasing the demands on working memory reduces the ability to ignore irrelevant stimuli.

One of the most important characteristics of working memory is its limited capacity Cowan (2001); Baddeley (2003). Working memory capacity helps to predict fluid intelligence and attentional control Engle *et al.* (1999a); Fukuda *et al.* (2010). For visual objects, this value has been estimated to be three or four items William (1974); Pashler (1988); Luck and Vogel (1997); Vogel *et al.* (2001). Studies that examine the visual working memory by sequentially presenting items have shown that the information is either completely stored or entirely forgotten William and Christie (1977); Smyth *et al.* (2005); Johnson and Miles (2008). Many studies have shown that as the number of objects to be stored in working memory increases, the precision gradually

decreases and it is worse for a sequential array of objects than a simultaneous array of objects Lecerf and de Ribaupierre (2005); Allen *et al.* (2006); Blalock and Clegg (2010).

An important question to examine is the effect of positive and negative information on visual working memory. Various studies have shown that emotional content has increased the chances of retaining the information for a long period Buchanan and Adolphs (2002); Dolan (2002); Hamann (2001). Different experiments conducted to examine the link between emotion and working memory by inducing mood in the participants have shown a change in cognitive task performance Elliman *et al.* (1997); Gray (2001); Spies *et al.* (1996). Spies *et al.* (1996); Cheng and Holyoak (1985) have demonstrated that negative mood hinders the performance on tests of problem-solving, working memory, and attention. This may be due to intrusive thoughts and worries which distract the participants from the task at hand Eysenck and Calvo (1992); Seibert and Ellis (1991). Individuals may be more likely to direct attention consciously toward emotional stimuli or to elaborate on emotional information because of its personal relevance Doerksen and Shimamura (2001); Heuer and Reisberg (1990). Depending on the task at hand, having additional emotional stimuli can ease, if the task-relevant information is processed, or weaken if task-irrelevant information is processed, the working memory capacity and performance of an individual. Perlstein *et al.* (2002) has shown that emotional content has hindered the performance on working memory tasks.

In the proposed study, I examine the effect of image captioning with description on the working memory of humans. I aim to understand the impact of positive and negative outcomes on working memory associated with a neutral image and understand

the impact on a long-term memory as well. I hypothesize that positive descriptions will be retained for a longer period compared to negative descriptions associated with the outcome of an image. I would also like to understand if additional information associated with the image helps the participants in retaining the image captions for a longer time in the working memory.

Chapter 3

IMAGE DATASET

Eight neutral images were selected such that each image could have a positive or negative outcome. All images comprise situations relatable to daily life. Each image is associated with two descriptions: a positive description, which is the result of a positive outcome of the situation shown in the image, and a negative description, which is the result of a negative outcome of the situation shown in the image.

Positive Description	Negative Description
Kiara and Sandy have been friends since they joined Infi Tech. Kiara works in the HR department while Sandy has been working with the security team at Infi Tech. It has been their routine to meet every day for a cup of coffee at the lounge. Today when they meet for coffee Kiara realized that Sandy is super excited to share something with her. Sandy shares that she has been promoted and has been offered a good increment for the next pay cycle. Both of them celebrate in the honor of Sandy's success.	Kiara and Sandy have been friends since they joined Infi Tech. Kiara works in the HR department while Sandy has been working with the security team at Infi Tech. It has been their routine to meet every day for a cup of coffee at the lounge. Today when they meet for coffee Kiara realized that Sandy is very disappointed. After multiple requests, Sandy shares that she has been moved to a different department due to layoffs in the company, and her bonus has been canceled to this change for the next pay cycle. Kiara advises her friend to find ways to get back to her old team.

Table 3.1: Positive And Negative Descriptions Of Image-1

Positive Description	Negative Description
It's been twenty long years since Mr. Harry and Mrs. Harry moved into this house. This house has been home and everything for twenty years. Their kids grew up here and every party for New Year's Eve was hosted in their backyard. Mr. Harry has been promoted to Senior Director at his company and he has been asked to move to a different location to take the position. The entire family is excited to start a new life at a new location.	It's been twenty long years since Mr. Harry and Mrs. Harry moved into this house. This house has been home and everything for twenty years. Their kids grew up here and every party for New Year's Eve was hosted in their backyard. Due to Covid, both of them lost their jobs and have been trying to fix things for the last 10 months. They couldn't find an alternative and so they are moving out of the house with all the memories they gathered.

Table 3.2: Positive And Negative Descriptions Of Image-2

Positive Description	Negative Description
James goes to bed after a very tiring walk with his puppy, Snoopy. He fell asleep immediately as his day was productive because he completed his assignment on time and scored the highest in his class in the previous test. He starts dreaming about being announced as the only student who scored a 100% in the final exam. He comes home to realize that his family gifts him another puppy for succeeding in his academics. He suddenly wakes up from his sleep to realize it was just a good dream.	James goes to bed after a very tiring walk with his puppy, Snoopy. He does not fall asleep easily as his day was really bad because neither he was able to complete his assignment on time nor his test results were impressive. He starts dreaming about being announced as the only student who failed in the final exam. He comes home with a heavy heart to realize that his puppy is missing as he forgot to lock the door before leaving for school. He suddenly wakes up from his sleep to realize it was just a bad dream and nothing more than that.

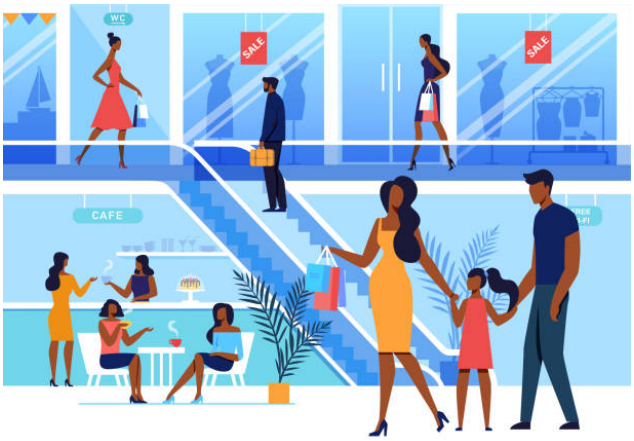
Table 3.3: Positive And Negative Descriptions Of Image-3

Positive Description	Negative Description
Sasha likes to explore cooking different dishes. She wishes to invite her friends home over this weekend for a get-together. So she begins to experiment with making a new dish by looking at the cookbook. Halfway through the cooking process, Sasha realized that the dish smells great and looks forward to tasting the dish. As she takes her first bite of the final dish, tears roll down her eyes realizing that it was one of the best dishes made by her to date. She is excited for her friends to try this and experience the same delight.	Sasha likes to explore cooking different dishes. She wishes to invite her friends home over this weekend for a get-together. So she begins to experiment with making a new dish by looking at the cookbook. Halfway through the cooking process, Sasha realized that the dish seemed incomplete and waited to taste the final outcome. As she takes her first bite of the final dish, she finds the dish is missing something essential and all the ingredients are not mixed well to taste good. She is disappointed to call her friends for dinner.

Table 3.4: Positive And Negative Descriptions Of Image-4

Positive Description	Negative Description
After graduating with Bachelors's in Computer Science, Samantha has been trying to get a job as a software engineer. She has been applying to a lot of companies but does not hear back anything positive. After waiting for two months, she finally gets a call for an interview from her dream company. She attends the interview with full confidence. She comes out of the interview hall appointed as the software developer for the upcoming new team.	After graduating with Bachelors's in Computer Science, Samantha has been trying to get a job as a software engineer. She has been applying to a lot of companies but does not hear back anything positive. After waiting for two months, she finally attends an interview. Worrying about the outcome, she does not answer the basic questions asked by the interviewer properly. She comes out of the interview hall with the disappointment of not getting a job.

Table 3.5: Positive And Negative Descriptions Of Image-5



Positive Description	Negative Description
Rita and her family come out to a mall to shop on the eve of her birthday coming next week. Rita is excited to see people and her parents calm her down to be careful not to get lost in the mall while shopping. The family gets into a shop to select clothes for Rita and seeing her favorite color, Rita runs to get them. Seeing Rita run, mom follows her. Mom stops her from getting into the crowd and warns Rita to be careful next time.	Rita and her family come out to a mall to shop on the eve of her birthday coming next week. Rita is excited to see people and her parents calm her down to be careful not to get lost in the mall while shopping. The family gets into a shop to select clothes for Rita and seeing her favorite color, Rita runs to get them. Seeing Rita run, mom follows her but due to a lot of people at the dress section, Rita gets lost and does not find her parents around her. She panics and searches for her parents. The shop executive finds Rita crying looking for her parents and announces it in the shop so that her parents can get her.

Table 3.6: Positive And Negative Descriptions Of Image-6

Positive Description	Negative Description
Robert is driving back home after a very long and tiring day at work. He keeps on getting office calls while driving back home as the project he is working on is in the final stage. He checks his phone to reply to a message while driving and realizes that he is at the zebra crossing. He immediately hits the car breaks seeing a woman walking in front of him crossing the road. He just missed dashing the woman and the baby with her. Robert realizes his mistake of using the phone while driving and never plans to repeat it.	Robert is driving back home after a very long and tiring day at work. He keeps on getting office calls while driving back home as the project he is working on is in the final stage. He checks his phone to reply to a message while driving and realizes that he is at the zebra crossing. Due to his high speed, he does not get to control his car and dashes into the women crossing the road. The whole incident leads to a big accident injuring him and the woman and her baby crossing the road.

Table 3.7: Positive And Negative Descriptions Of Image-7

Positive Description	Negative Description
Lisa is a senior year student. She comes back home after a tiring day at school. She has a habit of writing down her personal experiences for the day. After a long time, Lisa is happy to write down in her diary that she became the President of the Women's club at her school. It has been quite challenging work to campaign and propagating her agenda for the club. She is extremely happy that all her hard work has been paid and she landed in the position she wanted.	Lisa is a senior year student. She comes back home after a tiring day at school. She has a habit of writing down her personal experiences for the day. Lisa is disappointed as she loses the election for the President of the Womenś club at her school. It has been quite challenging work to campaign and propagates her agenda for the club. She has been working hard to get this position but she is heartbroken to lose the elections.

Table 3.8: Positive And Negative Descriptions Of Image-8

Chapter 4

EXPERIMENT

4.1 Participants

The total number of participants enrolled in this IRB-approved study was 65 undergraduate and graduate students from Arizona State University between ages 17-27. All the participants were well acquainted with the English language and had basic computer usage skills. I only used data from 50 participants for the analysis.

4.2 Procedure

The study consists of three rounds performed with a minimum gap of three days to a maximum gap of five days between rounds. For the first and second rounds, the experiment consisted of two parts. In the first part, an image was randomly selected from the set of eight images and was displayed for the subject to caption. In the second part of the round, the same image with a description, i.e, either positive or negative, was selected randomly and displayed. The participant captioned the image again after reading the description associated with it. Then a question was asked to check if the participant understood the description correctly or not. The participants had to finish the first part and second part for each image before moving on to the next image. The number of positive and negative descriptions for the images was equally balanced to maintain the consistency of the game. The only difference between the first and the second round was that the description in the first round was chosen at random for each participant whereas, in the second round, the description was opposite of the description displayed in the first round. For the third round, the

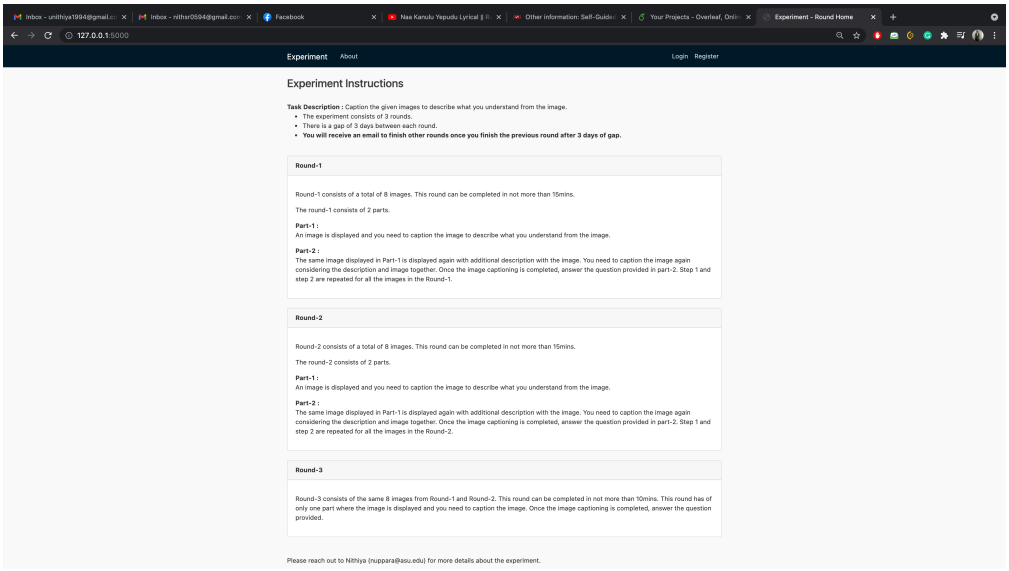


Figure 4.1: Instructions Page Of The Image Captioning Study.

participant is displayed the same images without any caption, which he or she saw in the previous rounds, and was asked to caption the image. The time taken to complete each round was captured using the software.

A website has been built to conduct the study to collect data from the participants. The home page of the website provides all the instructions for the participant to successfully complete the study. At the start of the study, the participant creates an account on the website with his information such as name, age, email, and password. The primary purpose of this account is to help the participant use the website for the further rounds of the study. To avoid participants to finish the rounds consecutively on the same day, the next rounds are locked and can only be accessed after three days from the completion of the previous round. I have also created alerts through email for the participants to come back and finish the study after three days after the previous round.

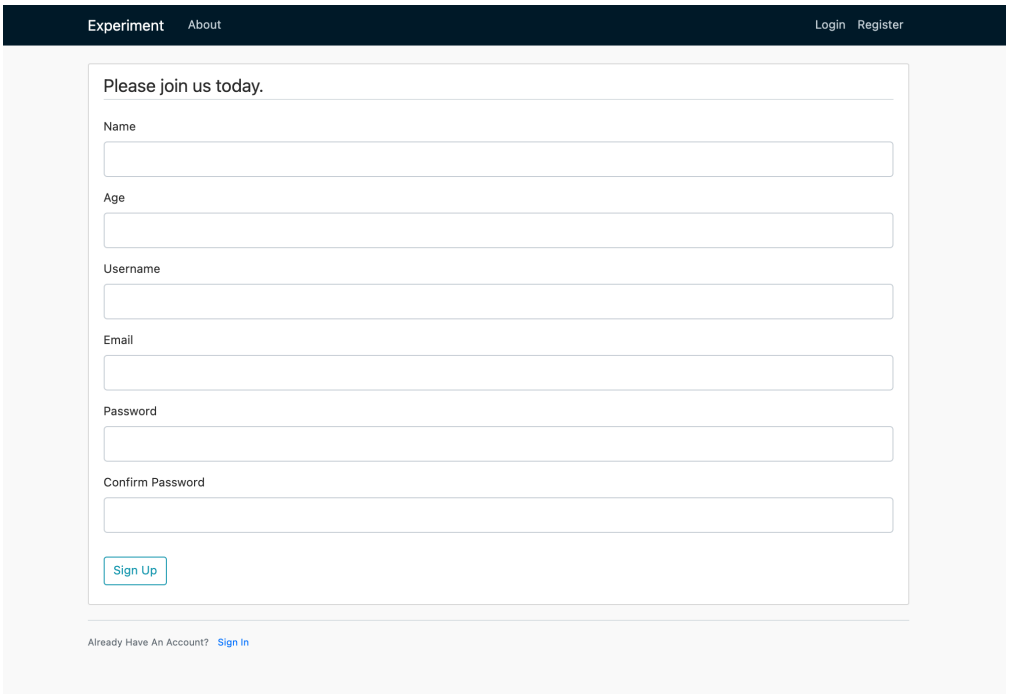


Figure 4.2: Registration Page Of The Image Captioning Study.

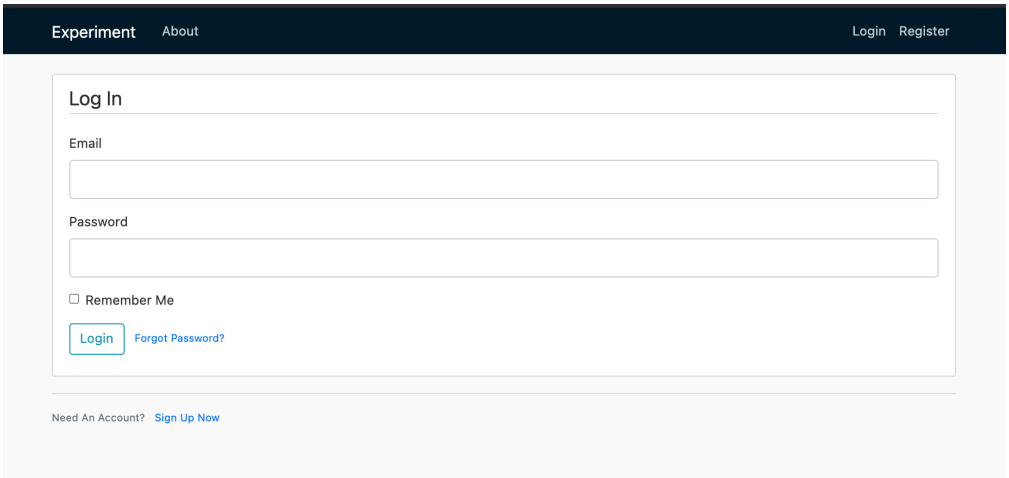


Figure 4.3: Login Page Of The Image Captioning Study.

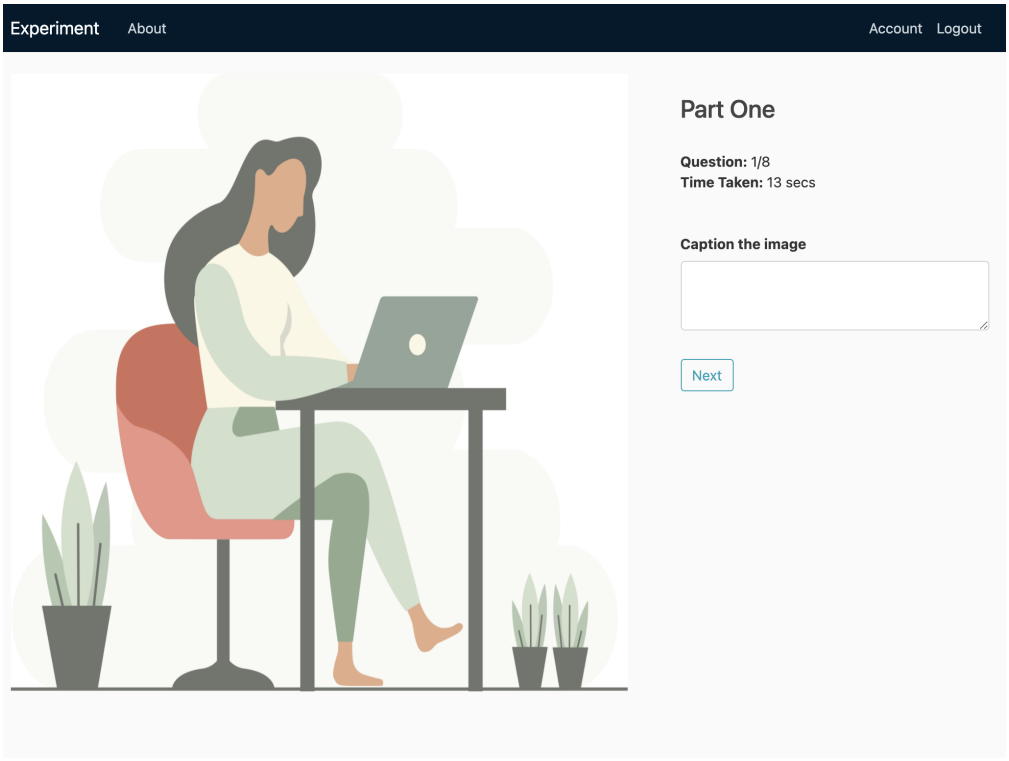


Figure 4.4: Part-1 Interface Of Round-1 And Round-2 Of The Image Captioning Study.

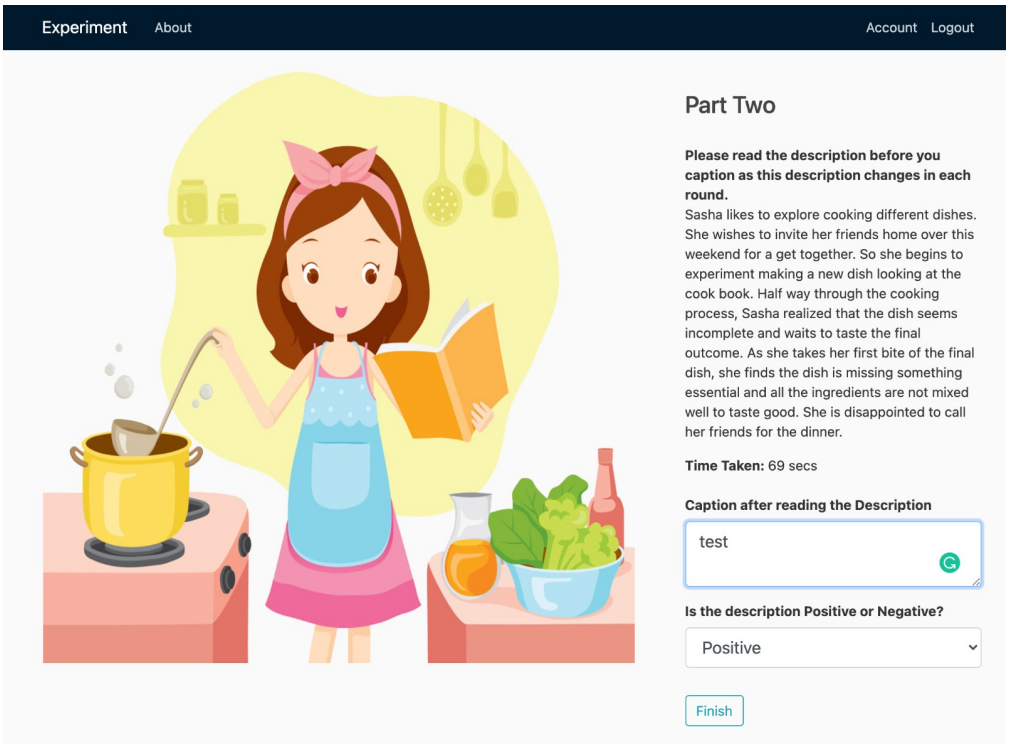


Figure 4.5: Part-2 Interface Of Round-1 And Round-2 Of The Image Captioning Study.

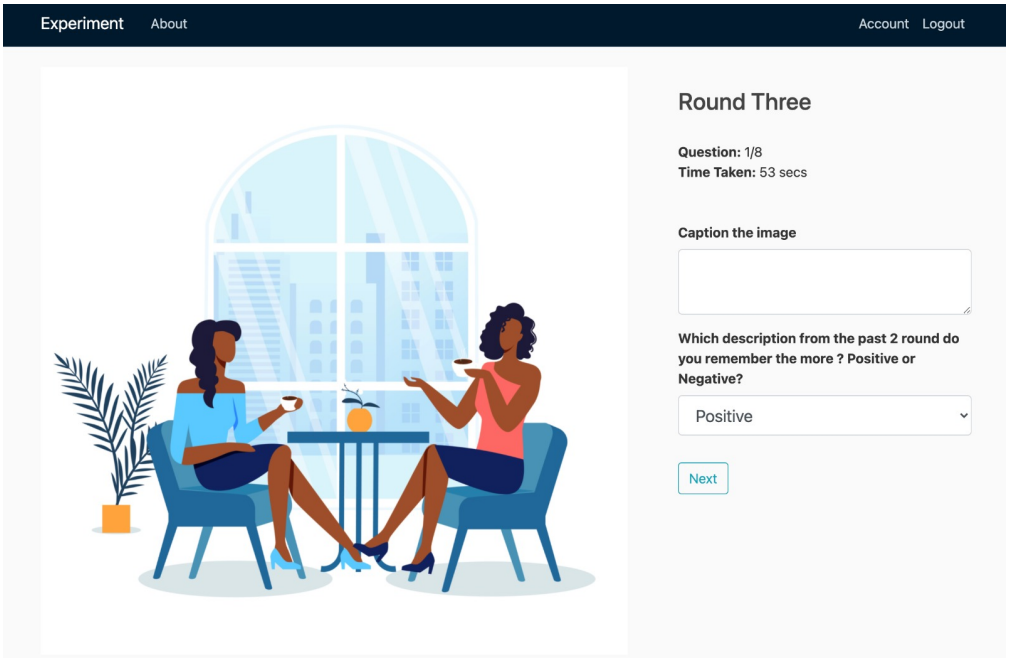


Figure 4.6: Round-3 Interface Of The Image Captioning Study.

4.3 Data Analysis

The total number of participants for the study was 65 but only 50 participants' data has been used for the analysis because either some of them didn't finish the 3 rounds or some of them didn't finish the rounds with the expected gap between each other.

Each subject inputs 5 captions for each of the 8 images. The first caption is named Round-1 Part-1 (R1-P1), the second caption is named Round-1 Part-2 (R1-P2), the third caption is named Round-2 Part-1 (R2-P1), the fourth caption is named as Round-2 Part-2 (R2-P2) and the fifth caption is named as Round-3 (R3).

I propose to analyze the captions to test for similarity and to evaluate sentiment. It is possible that two captions from the same user are similar in context but differ in language. To account for such cases, we used the HuggingFace BERT Sequence Classification pre-trained model Wolf *et al.* (2020) to identify contextual similarity. We also used the HuggingFace BERT Sequence Classification pre-trained model to identify the sentiment of the caption.

The main purpose of conducting different rounds of the experiment with regular intervals of time is to observe the recall span and retention span of the image captions when additional information like description is provided with the image. The possible combinations for this purpose considered are the ability to remember captions from R1-P2 in R2-P1, from R2-P1 in R3, and from R1-P1 in R3.

Chapter 5

RESULTS

5.1 Round-1 Part-1 (R1-P1) vs. Round-3 (R3)

Figure 5.1 shows that only 25% of the captions, i.e., 101, were contextually the same, and 75% of the captions, i.e., 299, were contextually different for R1-P1 and R3. Figure 5.2 shows the trend in the number of the participants who captioned the image contextually the same and different for each image from the image dataset. It is interesting to note that even after looking at the same image five times in total, and captioning it four times before R3, 75% of participants could not recollect the first caption they used to caption the image.

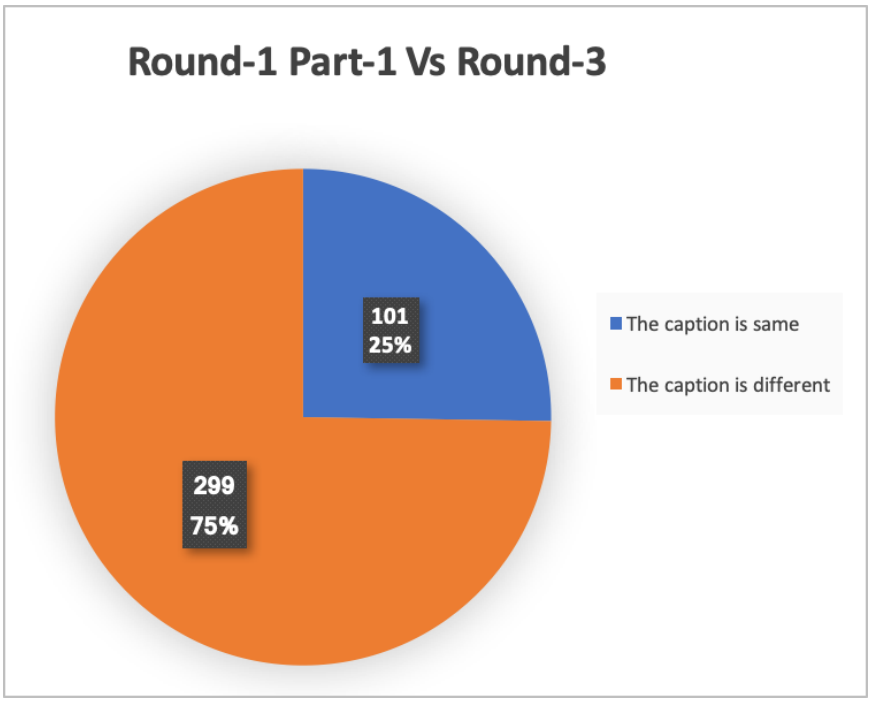


Figure 5.1: Distribution Of Image Captions From R1-P1 Which Are Contextually The Same As R3.

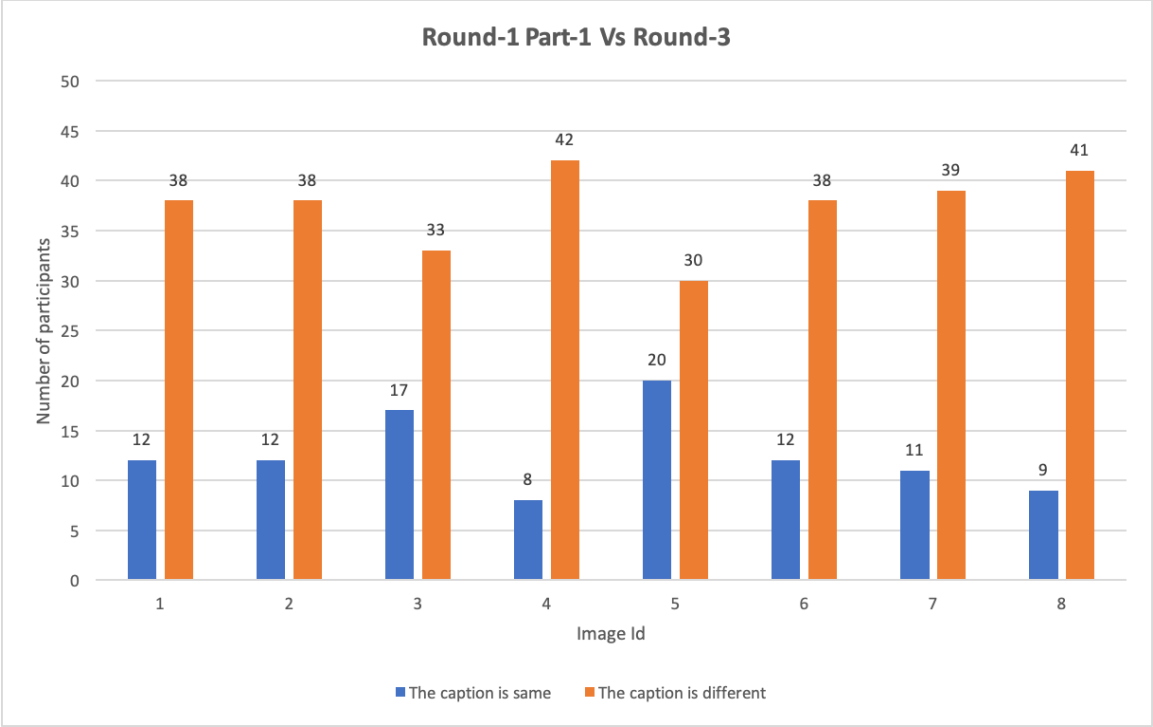


Figure 5.2: Trend In The Number Of Participants Who Captioned The Image In R3 Contextually The Same And Different As R1-P1.

5.2 Round-1 vs. Round-2 and Round-2 vs. Round-3

5.2.1 Round-1 (R1-P1 and R1-P2) vs. Round-2 Part-1 (R2-P1)

Figure 5.3 shows that of the participants who finished Round-1 (R1) and have seen the images twice, 36% of the participants, i.e., 143, captioned the image in R2-P1 contextually the same as in R1-P1; 13% of the participants, i.e., 54, captioned the image in R2-P1 contextually the same as in R1-P2; and the rest, 51%, of the participants, i.e., 203, captioned the image differently from the previous round. Figure 5.4 shows the trend in the number of the participants who captioned the image in R3 contextually the same as R1-P1, R1-P2, and the rest different for each image from the image dataset.

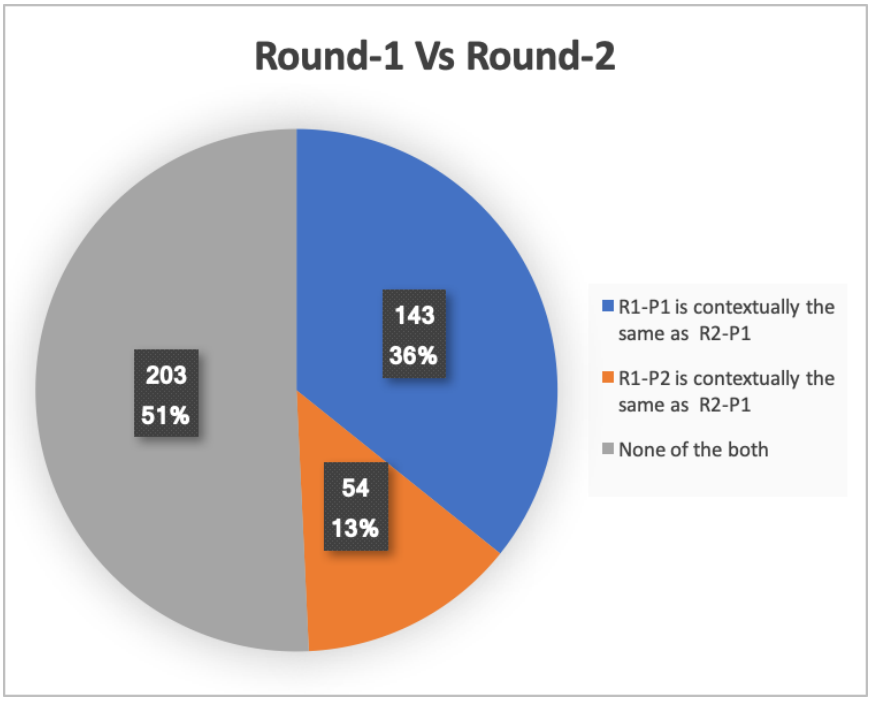


Figure 5.3: Distribution Of Image Captions From R1-P1 And R1-P2 Which Are Contextually The Same As R2-P1.

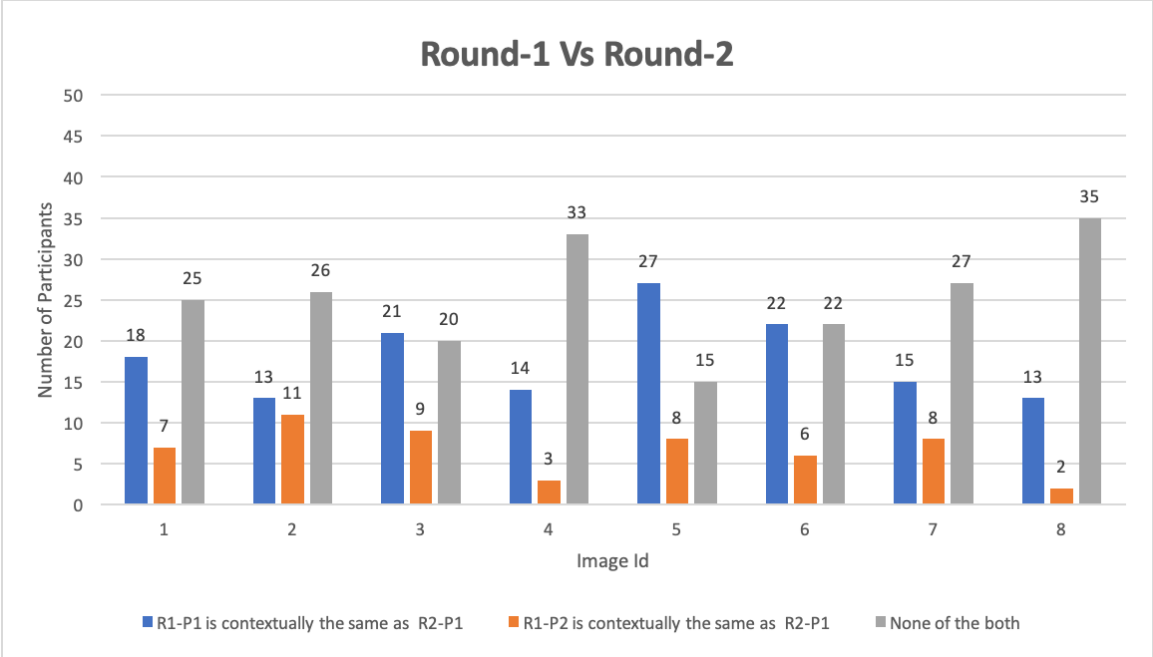


Figure 5.4: Trend In The Number Of Participants Who Captioned The Image In R2 Contextually The Same As R1-P1, R1-P2 And Not The Same.

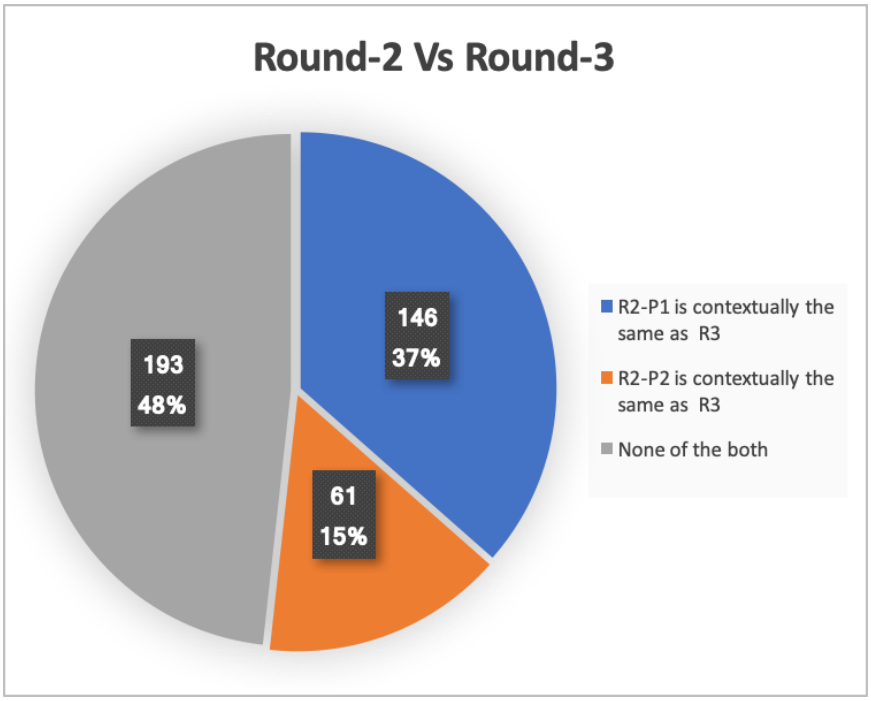


Figure 5.5: Distribution Of Image Captions From R2-P1 And R2-P2 Which Are Contextually The Same As R3.

5.2.2 Round-2 (R2-P1 and R2-P2) vs. Round-3 (R3)

Figure 5.5 shows that of the participants who completed Round-2 (R2) and have seen the images four times, 37% of the participants, i.e., 146, captioned the image in R3 contextually the same as in R2-P1; 15% of participants, i.e., 61, captioned the image in R3 contextually the same as R2-P2; and the rest, 49%, of participants, i.e., 193, captioned the image differently from the previous round. Figure 5.6 shows the trend in the number of participants who captioned the image in R3 contextually the same as in R2-P1, R2-P2, and none of both, i.e., the rest are different for each image from the image dataset.

An interesting observation is that even after seeing the image with extra information like the description associated with the image in R1 and R2, the majority of participants tend to remember the image caption they captioned in R1-P1 and R2-P2

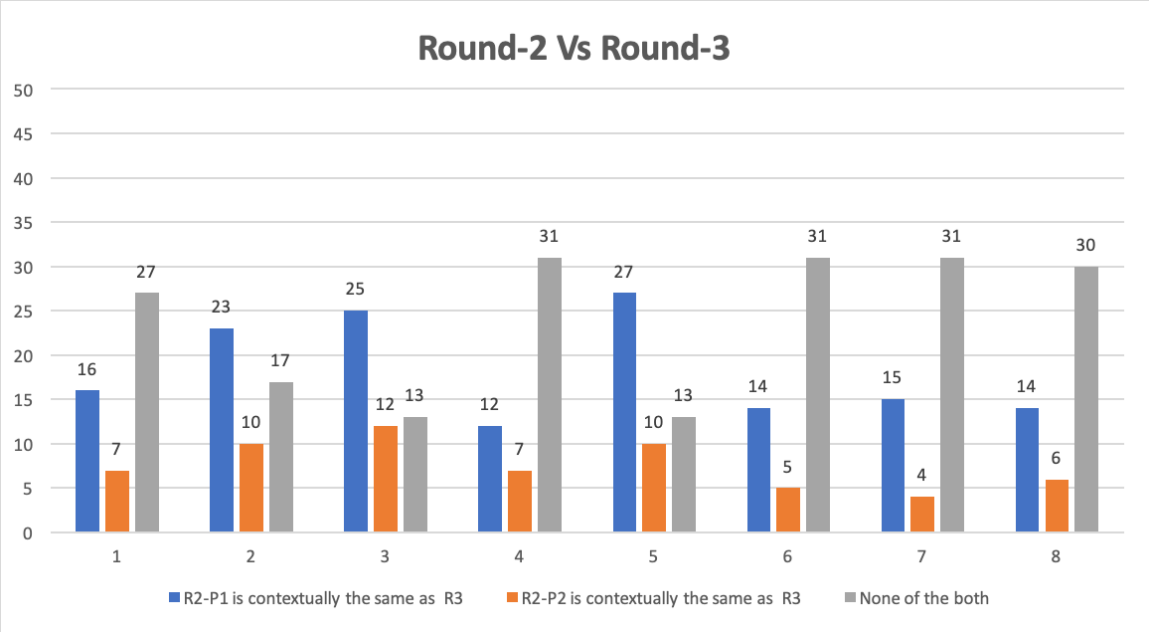


Figure 5.6: Trend In The Number Of The Participants Who Captioned The Image In R3 Contextually The Same As R2-P1, R2-P2 And Not The Same.

respectively.

5.3 Trends In Round-1 Part-2 And Round-2 Part-2

Figure 5.7 shows that among the total number of participants whose image caption from R1-P2 is contextually the same as R1-P1, 67% of participants, i.e., 36, captioned the image with respect to the positive description associated to it, and 33% of participants, i.e., 18, captioned the image with respect to the negative description associated to it.

Figure 5.8 shows that among the total number of participants whose caption from R2-P2 is contextually the same as in R3, 56% of the participants, i.e., 34, captioned the image with respect to the positive description associated to it, and 44% of the participants, i.e., 27, captioned the image with respect to the negative description associated to it.

It is interesting to note that out of the captions remembered by participants from

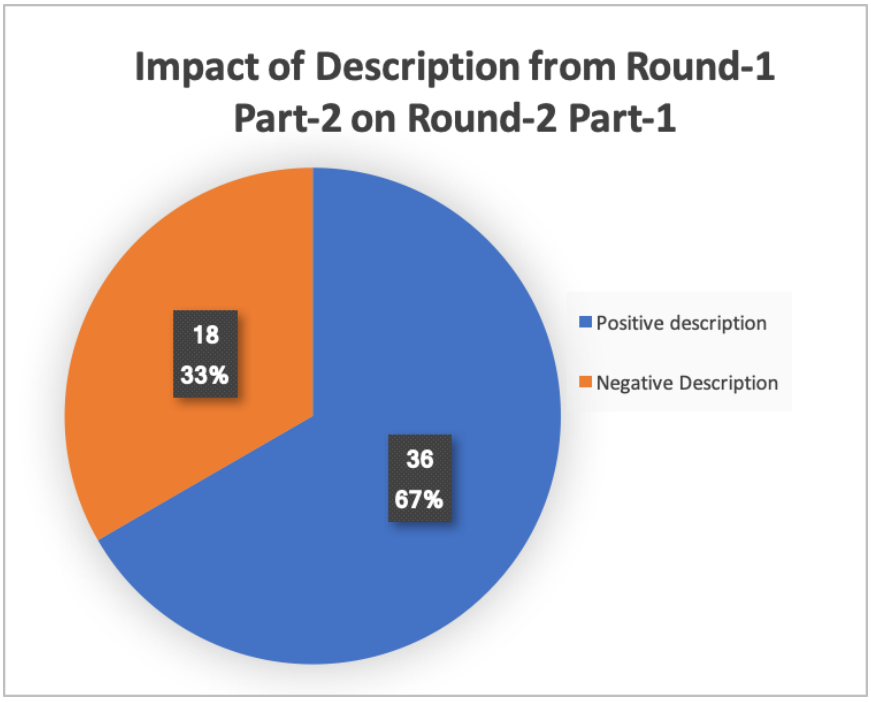


Figure 5.7: Impact Of Positive And Negative Description From R1-P2 On R2-P1.

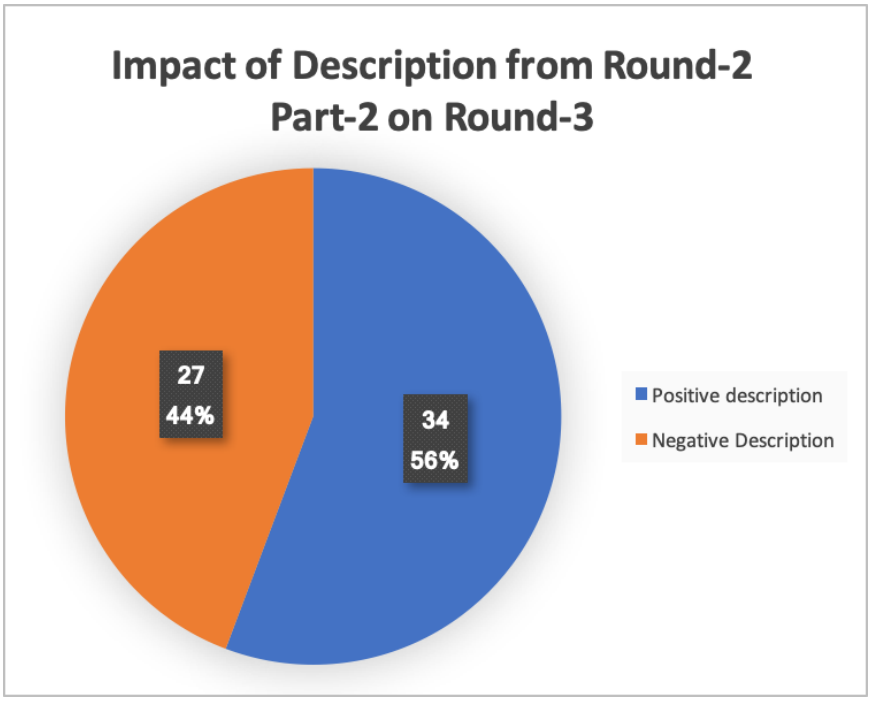


Figure 5.8: Impact Of Positive And Negative Description From R2-P2 On R3.

R1-P2 and R2-P2, positive descriptions tend to have more impact on participants, causing them to remember the image caption for a longer duration compared to negative descriptions.

Chapter 6

DISCUSSIONS

Only 25% of participants were able to recall the captions which they captioned for an image after 9-15 days. It is interesting that even though the visual working memory capacity of a human is considered to be three to four objects, participants tended to retain some of the information for 9-15 days. This may be due to the fact that some participants were able to relate the situations from the images leading them to correlate the image with one or more experiences from their past, consistent with Doerksen and Shimamura (2001); Heuer and Reisberg (1990). Due to this, even though the images were of no purpose to them, they tended to remember the captions for a long period, posing an interesting question to examine whether the image captions were saved to working memory or long-term memory. One other possible reason for retaining the image caption would be due to the additional information, i.e., the description, provided with the images. If providing description is a potential reason for participants to retain the information, it is fascinating to note that if the hypothesized reason behind remembering the image caption is that of the description, participants tended to recall the image caption which was captioned without the description.

When comparing the recall rate of the captions between the first and second rounds, and the second and third rounds, more than 50% of participants were able to recall the image caption from previous rounds. Out of the 50%, an average of 36% of the captions recalled were the image captions which were captioned without seeing the descriptions. This helps to understand that even after seeing extra description related to a given image, the first impression of the image made on participants has

more impact and a higher chance to be retained in the working memory than the caption which had been captioned after seeing the description. This also leads to the question that given an image without any description, why is it easy for a human to perceive the image than the description associated with it and relate it.

The primary purpose of using two different descriptions for an image was to understand the impact of the sentiment of the description on image captioning. As hypothesized, out of the participants who recalled the image caption with description in R2 and R3, an average of 60% of participants remembered the caption associated with the positive description rather than the negative description. It may be concluded that given two outcomes, one positive and the other negative, the human brain on average tends to remember and retain the positive information corresponding to the situation rather than the negative information. This also leads to an interesting question that the working memory capacity of a human tends to change with the sentiment of the objects associated with it, which is consistent with Buchanan and Adolphs (2002); Dolan (2002); Hamann (2001); Perlstein *et al.* (2002).

Chapter 7

LIMITATIONS

The inferences from this study are limited due to the following reasons:

1. There is no evidence of the mood of each participant while participating in the study. It is possible that the state of mind and mood may be a potential reason for the participant to remember some information for longer duration of time.

2. It has been discussed and decided by the committee and the chair members that the eight images in the image dataset are all neutral images and they can have a positive and negative outcome associated with it. It may be possible that the images may be interpreted differently by different people according to their knowledge and intellect.

3. The positive and negative description associated with the images has been written with the approval of the committee and chair members but the sentiment of the description can vary according to the discretion and interpretation of an individual.

Chapter 8

CONCLUSION & FUTURE WORKS

8.1 CONCLUSION

The dissertation tries to understand the importance of working memory on human tasks such as image captioning. It also introduces descriptions i.e, positive and negative descriptions, associated with the images to understand the impact of these descriptions on the image captioning task. In the first section, the dissertation outlines the goals and motivation behind the problem statement. The dissertation primarily creates the image dataset for designing the image captioning study. The dissertation then progresses by developing the image captioning game which should be completed at regular intervals of time by the participants.

Similar to many other academic works, the research ideas presented in this thesis ask more questions than answers. Some of the questions are: How can human retain information in the working memory for a time period of 9 to 15 days? Is it because the information perceived from the images had already been in long-term memory in some other form or is it because the human being was correlating the image with his past memories?

Finally, the findings from the results highlight that participants tend to retain information for longer periods than the expected duration for working memory, which may be because participants were able to relate the images with their everyday life situations. Figure 5.7 and Figure 5.8 give insight that the positive description enabled participants to retain and recall more information than the negative description associated with the image. The most interesting outcome of the study can be summa-

rized as given a situation with positive and negative information, the human brain is aligned towards positive information over negative information. This can be a useful insight in designing future studies and collecting more data to dwell deep into understanding the additional reasons behind this human behavior. Even though there are some limitations to this study, the results contribute to the growing research on working memory and image captioning of humans.

8.2 FUTURE WORKS

- The study can be extended to analyze the sentiment of the image captions from the current data.
- The study can be extended to incorporate eye tracking to understand the point of interest of the participants. This will help us to understand in detail the reasons behind the information being retained in the working memory for more time than expected.
- The study can also incorporate a survey to record the mood and state of mind of the participant which can be analyzed to see any correlation between them to the results from the image captioning study.

BIBLIOGRAPHY

Allen, R., A. Baddeley and G. Hitch, "Is the binding of visual features in working memory resource-demanding?", Journal of experimental psychology. General **135**, 298–313 (2006).

Alvarez, G. A. and P. Cavanagh, "The capacity of visual short-term memory is set both by visual information load and by number of objects", Psychological science **15**, 2, 106–111 (2004).

Ashkenazi, S., M. Rosenberg-Lee, A. W. Metcalfe, A. G. Swigart and V. Menon, "Visuo–spatial working memory is an important source of domain-general vulnerability in the development of arithmetic cognition", Neuropsychologia **51**, 11, 2305–2317 (2013).

Atkinson, R. C. and R. M. Shiffrin, "Human memory: A proposed system and its control processes", in "Psychology of learning and motivation", vol. 2, pp. 89–195 (Elsevier, 1968).

Baars, B. J. and S. Franklin, "How conscious experience and working memory interact", Trends in cognitive sciences **7**, 4, 166–172 (2003).

Baddeley, A., "Baddeley a. working memory: looking back and looking forward. nat rev neurosci 4: 829-839", Nature reviews. Neuroscience **4**, 829–39 (2003).

Baddeley, A., "Working memory", Current Biology **20**, 4, R136–R140, URL `https://www.sciencedirect.com/science/article/pii/S0960982209021332` (2010).

Baddeley, A., "Working memory: Theories, models, and controversies", Annual Review of Psychology **63**, 1, 1–29, pMID: 21961947 (2012).

Baddeley, A. and G. Hitch, *Working memory*, vol. 8, pp. 47–90 (Academic Press, 1974).

Blalock, L. and B. Clegg, "Encoding and representation of simultaneous and sequential arrays in visuospatial working memory", Quarterly journal of experimental psychology (2006) **63**, 856–62 (2010).

Broadbent, D. E., *Perception and communication* (Pergamon Press, 1958).

Buchanan, T. and R. Adolphs, "The role of the human amygdala in emotional modulation of long-term declarative memory", Advances in Consciousness Research **44**, 9–34 (2002).

Callicott, J. H., V. S. Mattay, A. Bertolino, K. Finn, R. Coppola, J. A. Frank, T. E. Goldberg and D. R. Weinberger, "Physiological characteristics of capacity constraints in working memory as revealed by functional mri", Cerebral cortex **9**, 1, 20–26 (1999).

Chai, W. J., A. I. Abd Hamid and J. M. Abdullah, "Working memory from the psychological and neurosciences perspectives: A review", Frontiers in Psychology **9**, 401, URL https://www.frontiersin.org/article/10.3389/fpsyg.2018.00401 (2018).

Chelonis, J. J., J. L. Daniels-Shaw, D. J. Blake and M. G. Paule, "Developmental aspects of delayed matching-to-sample task performance in children", Neurotoxicology and Teratology **22**, 5, 683–694 (2000).

Cheng, P. and K. Holyoak, "Pragmatic reasoning schemas", Cognitive psychology **17**, 391–416 (1985).

Conklin, H. M., M. Luciana, C. J. Hooper and R. S. Yarger, "Working memory performance in typically developing children and adolescents: Behavioral evidence of protracted frontal lobe development", Developmental neuropsychology **31**, 1, 103–128 (2007).

Cowan, N., "The magical number 4 in short-term memory: A reconsideration of mental storage capacity", Behavioral and Brain Sciences **24**, 1, 87–114 (2001).

Cowan, N., "What are the differences between long-term, short-term, and working memory?", Progress in brain research **169**, 323–38 (2008).

Daneman, M. and P. A. Carpenter, "Individual differences in working memory and reading", Journal of Verbal Learning and Verbal Behavior **19**, 4, 450–466, URL https://www.sciencedirect.com/science/article/pii/S0022537180903126 (1980).

Daneman, M. and P. Merikle, "Working memory and language comprehension: A meta-analysis", Psychonomic bulletin & review **3**, 422–433 (1996).

Davelaar, E., "Short-term memory as a working memory control process", Frontiers in Psychology **4**, 13, URL https://www.frontiersin.org/article/10.3389/fpsyg.2013.00013 (2013).

de Fockert, J. W., G. Rees, C. D. Frith and N. Lavie, "The role of working memory in visual selective attention", Science **291**, 5509, 1803–1806 (2001).

D'Esposito, M. and B. R. Postle, "The cognitive neuroscience of working memory", Annual Review of Psychology **66**, 1, 115–142, pMID: 25251486 (2015).

Doerksen, S. and A. Shimamura, "Source memory enhancement for emotional words", Emotion (Washington, D.C.) **1**, 5–11 (2001).

Dolan, R., "Emotion, cognition, and behavior", Science (New York, N.Y.) **298**, 1191–4 (2002).

Downing, P. E., "Interactions between visual working memory and selective attention", Psychological science **11**, 6, 467–473 (2000).

Elliman, N., M. Green, P. Rogers and G. Finch, "Processing-efficiency theory and the working-memory system: Impairments associated with sub-clinical anxiety", Personality and Individual Differences **23**, 31–35 (1997).

Engle, R., S. Tuholski, J. Laughlin and A. Conway, "Working memory, short-term memory and general fluid intelligence: A latent variable approach", Journal of Experimental Psychology: General **130**, 169–183 (1999a).

Engle, R., S. W. Tuholski, J. Laughlin and A. Conway, "Working memory, short-term memory, and general fluid intelligence: a latent-variable approach.", Journal of experimental psychology. General **128 3**, 309–331 (1999b).

Eysenck, M. and M. Calvo, "Anxiety and performance: The processing efficiency theory", Cognition & Emotion - COGNITION EMOTION **6**, 409–434 (1992).

Fukuda, K., E. Vogel, U. Mayr and E. Awh, "Quantity, not quality: The relationship between fluid intelligence and working memory capacity", Psychonomic bulletin & review **17**, 673–9 (2010).

Gathercole, S. E., S. J. Pickering, B. Ambridge and H. Wearing, "The structure of working memory from 4 to 15 years of age.", Developmental psychology **40**, 2, 177 (2004).

Gray, J., "Emotional modulation of cognitive control: Approach-withdrawal states double dissociate spatial from verbal 2-back task performance", Journal of experimental psychology. General **130**, 436–52 (2001).

Hamann, S., "Cognitive and neural mechanisms of emotional memory", Trends in cognitive sciences **5**, 394–400 (2001).

Harden, L., *A review of research on working memory and its importance in education of the deaf*, Ph.D. thesis, Program in Audiology and Communication Sciences, Washington University, URL `http://digitalcommons.wustl.edu/pacs_capstones/627` (2011).

Heuer, F. and D. Reisberg, "Vivid memories of emotional events: The accuracy of remembered minutiae", Memory & Cognition **18**, 496–506 (1990).

Hitch, G. J., J. N. Towse and U. Hutton, "What limits children's working memory span? theoretical accounts and applications for scholastic development.", Journal of Experimental Psychology: General **130**, 2, 184 (2001).

Jaeggi, S. M., R. Seewer, A. C. Nirkko, D. Eckstein, G. Schroth, R. Groner and K. Gutbrod, "Does excessive memory load attenuate activation in the prefrontal cortex? load-dependent processing in single and dual tasks: functional magnetic resonance imaging study", NeuroImage **19**, 2, 210–225 (2003).

Jansma, J., N. Ramsey, N. Van Der Wee and R. Kahn, "Working memory capacity in schizophrenia: a parametric fmri study", Schizophrenia research **68**, 2-3, 159–171 (2004).

Johnson, A. and C. Miles, "Serial position effects in 2-alternative forced choice recognition: Functional equivalence across visual and auditory modalities", Memory (Hove, England) **17**, 84–91 (2008).

Just, M. and P. A. Carpenter, "A capacity theory of comprehension: Individual differences in working memory", URL `https://kilthub.cmu.edu/articles/journal_contribution/A_capacity_theory_of_comprehension_Individual_differences_in_working_memory/6613067/1` (2018).

Kim, C., J. K. Kroger, V. D. Calhoun and V. P. Clark, "The role of the frontopolar cortex in manipulation of integrated information in working memory", Neuroscience letters **595**, 25–29 (2015).

Kirschen, M., S. H. A. Chen, P. K. Schraedley-Desmond and J. Desmond, "Load- and practice-dependent increases in cerebro-cerebellar activation in verbal working memory: an fmri study", NeuroImage **24**, 462–472 (2005).

Kyllonen, P. C. and R. E. Christal, "Reasoning ability is (little more than) working-memory capacity?!", Intelligence **14**, 4, 389–433, URL `https://www.sciencedirect.com/science/article/pii/S0160289605800121` (1990).

Lavie, N., "Distracted and confused?: Selective attention under load", Trends in cognitive sciences **9**, 2, 75–82 (2005).

Lavie, N., A. Hirst, J. W. De Fockert and E. Viding, "Load theory of selective attention and cognitive control.", Journal of experimental psychology: General **133**, 3, 339 (2004).

Lecerf, T. and A. de Ribaupierre, "Recognition in a visuospatial memory task: The effect of presentation", European Journal of Cognitive Psychology - EUR J COGN PSYCHOL **17**, 47–75 (2005).

Logie, R. H. and R. Logie, *Visuo-spatial working memory* (Psychology Press, 1995).

Luck, S. and E. Vogel, "The capacity of visual working memory for features and conjunctions", Nature **390**, 279–81 (1997).

MacDonald, M., *Your Brain: The Missing Manual: The Missing Manual* (O'Reilly Media, 2008).

Miller, G., E. Galanter and K. Pribram, "Plans and the structure of behavior", The American Journal of Psychology **75** (1960).

Oberauer, K. and N. Cowan, "Working memory capacity: (2005)", Experimental Psychology - EXP PSYCHOL **54**, 245–246 (2005).

Pashler, H., "Familiarity and visual change detection", Perception & psychophysics **44**, 369–78 (1988).

Perlstein, W., T. Elbert and V. Stenger, "Perlstein wm, elbert t, stenger va. dissociation in human prefrontal cortex of affective influences on working memory-related activity. proc natl acad sci usa 99: 1736-1741", Proceedings of the National Academy of Sciences of the United States of America **99**, 1736–41 (2002).

Rypma, B., V. Prabhakaran, J. E. Desmond, G. H. Glover and J. D. Gabrieli, "Load-dependent roles of frontal brain regions in the maintenance of working memory", Neuroimage **9**, 2, 216–226 (1999).

Seibert, P. and H. Ellis, "Irrelevant thoughts, emotional mood states, and cognitive task performance", Memory & cognition **19**, 507–13 (1991).

Smyth, M., D. Hay, G. Hitch and N. Horton, "Serial position memory in the visual—spatial domain: Reconstructing sequences of unfamiliar faces", The Quarterly journal of experimental psychology. A, Human experimental psychology **58**, 909–30 (2005).

Spies, K., F. Hesse and C. Hummitzsch, "Mood and capacity in baddeley's model of human memory.", Zeitschrift für Psychologie mit Zeitschrift für angewandte Psychologie **204**, 367–381 (1996).

Thomason, M., E. A. Race, B. Burrows, S. Whitfield-Gabrieli, G. Glover and J. Gabrieli, "Development of spatial and verbal working memory capacity in the human brain", Journal of Cognitive Neuroscience **21**, 316–332 (2009).

Towse, J., G. Hitch and F. Hutton, "What limits children's working memory span? theoretical accounts and applications for scholastic development.", Journal of Experimental Psychology: General **130**, 2, 184–198, towse co-designed experiments. Towse analysed data. Towse co-wrote manuscript. Towse was PI on grant (ESRC R000236113) that funded the research. Towse presented the results at the ESCoP conference (2001). RAE$_{i}mport_{t}ype$: $JournalarticleRAE_{u}oa_{t}ype$: $Psychology$(2001).

Veltman, D. J., S. A. Rombouts and R. J. Dolan, "Maintenance versus manipulation in verbal working memory revisited: an fmri study", Neuroimage **18**, 2, 247–256 (2003).

Vogel, E., G. Woodman and S. Luck, "Storage of features, conjunctions, and objects in visual working memory", Journal of experimental psychology. Human perception and performance **27**, 92–114 (2001).

William, P., "Phillips, w. a. on the distinction between sensory storage and short-term visual memory. percept. psychophys. 16, 283-290", Perception & Psychophysics **16**, 283–290 (1974).

William, P. and D. Christie, "Components of visual memory", Quarterly Journal of Experimental Psychology - QUART J EXP PSYCHOL **29**, 117–133 (1977).

Wolf, T., L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite,

J. Plu, C. Xu, T. L. Scao, S. Gugger, M. Drame, Q. Lhoest and A. M. Rush, "Transformers: State-of-the-art natural language processing", in "Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations", pp. 38–45 (Association for Computational Linguistics, Online, 2020), URL https://www.aclweb.org/anthology/2020.emnlp-demos.6.

Zarahn, E., B. Rakitin, D. Abela, J. Flynn and Y. Stern, "Positive Evidence against Human Hippocampal Involvement in Working Memory Maintenance of Familiar Stimuli", Cerebral Cortex **15**, 3, 303–316, URL https://doi.org/10.1093/cercor/bhh132 (2004).

APPENDIX A

UNIVERSITY APPROVAL FOR HUMAN TESTING

EXEMPTION GRANTED

Troy McDaniel
IAFSE-PS: Polytechnic Engineering Programs (EGR)
480/727-1063
Troy.McDaniel@asu.edu

Dear Troy McDaniel:

On 3/22/2021 the ASU IRB reviewed the following protocol:

Type of Review:	Initial Study
Title:	Understanding the effect of image captioning on working memory.
Investigator:	Troy McDaniel
IRB ID:	STUDY00013587
Funding:	None
Grant Title:	None
Grant ID:	None
Documents Reviewed:	• Consent Form, Category: Consent Form; • Instructions, Category: Participant materials (specific directions for them); • IRB Social Behavior , Category: IRB Protocol; • Questions, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • Recruitment Method (Email), Category: Recruitment Materials;

The IRB determined that the protocol is considered exempt pursuant to Federal Regulations 45CFR46 (2) Tests, surveys, interviews, or observation on 3/22/2021.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

If any changes are made to the study, the IRB must be notified at research.integrity@asu.edu to determine if additional reviews/approvals are required. Changes may include but not limited to revisions to data collection, survey and/or interview questions, and vulnerable populations, etc.

Sincerely,

IRB Administrator

cc: Nithiya Uppara
Nithiya Uppara
Hemanth Kumar Demakethepalli Venkateswara

APPENDIX B

PERMISSION STATEMENTS FROM CO-AUTHORS

Permission for including co-authored material in this dissertation was obtained from co-authors, Dr. Troy McDaniel and Dr. Hemanth Venkateswara.